A multiple sequence alignment (MSA) with $N$ sequences $x^n = (x_1, \ldots, x_L)$ of length $L$ can be represented as a data matrix of $L$ categorical variables with 21 states (20 amino acids and gap symbol). In order to visualize the substructure of alignments we applied a principal component analysis (PCA) analysis on the MSA. The protein sequences of the MSA were converted to binary vectors using one-hot encoding and PCA was then performed on the transformed data matrix. Sequences were projected onto the first two principal components.